



Skill networks and measures of complex human capital

Katharine A. Anderson^{a,1}

^aTepper School of Business, Carnegie Mellon University, Pittsburgh, PA 15213

Edited by Matthew O. Jackson, Stanford University, Stanford, CA, and approved October 24, 2017 (received for review April 20, 2017)

We propose a network-based method for measuring worker skills. We illustrate the method using data from an online freelance website. Using the tools of network analysis, we divide skills into endogenous categories based on their relationship with other skills in the market. Workers who specialize in these different areas earn dramatically different wages. We then show that, in this market, network-based measures of human capital provide additional insight into wages beyond traditional measures. In particular, we show that workers with diverse skills earn higher wages than those with more specialized skills. Moreover, we can distinguish between two different types of workers benefiting from skill diversity: jacks-of-all-trades, whose skills can be applied independently on a wide range of jobs, and synergistic workers, whose skills are useful in combination and fill a hole in the labor market. On average, workers whose skills are synergistic earn more than jacks-of-all-trades.

labor economics | human capital | skill diversity | network analysis

The relationship between worker skills and wages is a problem of tremendous economic interest, making it critical to have effective measures of the skills, knowledge, and experience that a worker brings to production: a bundle of worker characteristics that economists refer to as human capital. Traditionally, human capital measures either divide workers into broad categories (e.g., laborers and management) or count years of experience, training, or education (1). However, treating skills as interchangeable removes some of the richness of human capital: workers' skills are clearly heterogeneous and multidimensional as are the skills required for jobs. A considerable body of literature shows the importance of skill diversity, specialization, and recombination in problem-solving and knowledge generation (2–8). This plus continued growth in knowledge-based industry (9) have generated interest in more nuanced measures of human capital (10–19).

Determining the relationship between wages and factors like skill diversity requires us to not only look at a worker's individual skills but also, her skill combinations. However, considering skills in combination makes measuring human capital much more difficult (20). Some skills (e.g., programing and user interface design) are synergistic, meaning that the combination is more valuable than the sum of its parts: each skill enhances the effectiveness of the other. Other skills (e.g., programing and Russian translation) are no more valuable together than they are individually. On the supply side, some skills (e.g., programing and management skills) are quite common individually but extremely rare in combination. Taken together, these factors mean that the value of an additional skill will depend on the skills that the worker already has (16, 21).

Here, we propose a network-based framework for the characterization of human capital that complements existing notions of human capital and production. Given a pool of workers with multidimensional skill baskets, we construct a network in which skills are nodes and two skills are connected by a link if a worker has both. Links are weighted according to how often the two skills co-occur. We construct a similar network using the skill sets required to perform different jobs, wherein two skills are connected if they are required by an employer in combination. Together, these two networks provide a more complete picture

of the supply and demand for human capital in a particular job market. Most importantly, they suggest a number of measures of human capital that account for both the relationships between skills and the context dependency of human capital.

We then use data from an online freelance labor market as an illustration of the method. Using information drawn from worker profiles and employer job advertisements, we construct a human capital network and several network-based measures of worker skills. We use an algorithmic method to split the human capital networks into clusters of closely related skills, providing an entirely endogenous categorization of skills. There is considerable variation in wages between workers specializing in these different skill categories. Workers with more diverse skills tend to earn higher wages than specialists. We then compare the skill measures on the supply and demand sides of the market to show that workers with diverse skills fall into two categories: those who exploit gaps in the market tend to earn higher wages than “jacks-of-all-trades,” who use their skills for multiple different jobs. Finally, we illustrate the value of this approach by showing that our network-based human capital measures explain variation in worker wages, even after controlling for individual skills.

Methods: Constructing a Human Capital Network

Network science provides a means of making sense of the relationships between skills in the labor market. Here, we construct two different networks: one representing the skills that workers have and the other representing the skills that employers require. Nodes in these networks are skills present in the labor market. On the workers' side, two skills are connected by a link if the same worker has both. On the employers' side, two skills are connected if they are required for the same job. We will call these networks the worker (supply) side and the employer (demand) side human capital networks.

Significance

The relationship between worker human capital and wages is a question of considerable economic interest. Skills are usually characterized using a one-dimensional measure, such as years of training. However, in knowledge-based production, the interaction between a worker's skills is also important. Here, we propose a network-based method for characterizing worker skill sets. We construct a human capital network, wherein nodes are skills and two skills are connected if a worker has both or both are required for the same job. We then illustrate the method by analyzing an online freelance labor market, showing that workers with diverse skills earn higher wages and that those who use their diverse skills in combination earn the highest wages of all.

Author contributions: K.A.A. designed research, performed research, analyzed data, and wrote the paper.

The author declares no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹Email: andersok@andrew.cmu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1706597114/-DCSupplemental.

More formally, let $I = \{1, 2, \dots, N\}$ be a pool of workers, each endowed with a skill set $A_i = \{s_1, s_2, \dots, s_k\}$. Let $S_W = \bigcup_{i \in I} \{A_i\}$ denote the set of all skills possessed by workers, with $|S_W| = M_W$. Let $A = \{A_1, A_2, \dots, A_N\}$ denote the set of all worker skill sets. Let n_i be the number of skill sets containing skill s_i , and let n_{ij} be the number of skill sets containing both s_i and s_j . (Note that we have assumed that skills are binary. Our data do not allow us to consider the ability level of an individual in a particular skill, and therefore, we do not consider it explicitly. However, the intensive margin could easily be incorporated into either the link weights or the measures derived from the network.) In a worker human capital network, $g(A)$, the nodes are the skills in the set S_W , and two skills, s_j and s_k , are connected if $s_j, s_k \in A_i$ for some $A_i \in A$. This network can be represented as an $M_W \times M_W$ matrix, where $g_{jk} = w_{jk}$ if $s_j, s_k \in A_i$ for some $A_i \in A$ and is zero otherwise. The value w_{jk} is the weight on the link between skills i and j .

Let $K = \{1, 2, \dots, F\}$ be the set of vacancies in the labor market. Let $B_f = \{s_1, s_2, \dots, s_k\}$ be the set of skills required for job f , and let $B = \{B_1, B_2, \dots, B_F\}$. Let $S_J = \bigcup_{i \in I} \{B_i\}$ denote the set of all skills requested by all employers in the market with $|S_J| = M_J$. Then, one can construct a network similar to that above, $g(B)$, an $M_J \times M_J$ matrix where $g_{jk} = w_{jk}$ if $s_j, s_k \in B_i$ for some $B_i \in B$ and is $g_{jk} = 0$ otherwise. Note that this network might be much different from the human capital possessed on the supply side.

We will weight each link in the network to reflect how closely related the two skills are in the labor market. Here, our weights will be a modification of conditional probability, which we will call skill similarity weights (alternative weighting schemes are discussed in *SI Appendix*): $w_{ij}^{sim} = P(s_i | s_j) = n_{ij} / n_j$, where n_i and n_j are the numbers of workers who have skills i and j , respectively, and $n_j < n_i$. Skill similarity weights have three properties that are desirable in this context.

- i) If skill A never co-occurs with skill B ($A \cap B = \emptyset$), then the link weight is zero.
- ii) If skill A always occurs with skill B ($A \subseteq B$), then the link weight is one.
- iii) The weight between two skills is strictly increasing as they co-occur more frequently.

Illustration of Network-Based Measures of Human Capital

The worker and job networks effectively summarize the human capital in a particular job market. A major advantage of this method is that it is context-dependent—a human capital network is constructed using data from a particular labor pool, meaning that a worker's human capital measures will also be context-dependent, which reflects differences in how her skills are valued in different markets (17). We will now show this method in a particular context: an online freelance labor market called UpWork. (These data both are publically available and do not require interaction with any individual. Thus, it does not qualify as human subjects research according to the Institutional Review Board at Carnegie Mellon University.)

Data. The world of online freelancing is a natural data source for this method. Online freelance labor has been a growing part of the economy fueled by a combination of better technology and evolving attitudes toward career change, and that growth is expected to continue (22). As a result, UpWork and associated markets (e.g., Elance and ODesk) have been the subject of much recent study (23–27).

In the UpWork market, workers apply for jobs, and employers hire and pay workers through the site. The information on a worker's profile includes a list of her skills and a list of previous jobs with associated hourly wages. Job postings contain information about the job, including a list of required skills. The

job opportunities range from small tasks, such as data entry and software testing, to large-scale projects, like application development and website design.

We use a sample of 26,046 worker profiles and 365,561 job listings collected over a period of 3 mo between November 2013 and January 2014. We use this full population to construct our skill networks, because all of the workers are visible to employers and thus, part of the labor market; 18,283 workers have a wage history on the site. A worker's average hourly wage is calculated from her wage history, which both is publicly visible and cannot be altered (26). While both hourly and flat rate jobs are listed, we only consider hourly jobs in calculating the wage rate, because we do not observe the hours worked on fixed price jobs. The average worker on the site makes \$16.74/h, has worked 765 h on the site, and lists six to seven skills. The distribution of workers' hourly wages shows typical inequality (*SI Appendix* has summary statistics).

The worker profiles include 2,197 different skills, and the job postings include 2,447. Workers and employers must choose their listed skills from the site's database of allowable skills. Adding a skill to this database requires a petition by a worker/employer and is only granted if the skill is not redundant. This eliminates any ambiguity in skills caused by spelling errors or synonymous entries and makes the data ideal for this application (23). We drop skills that occur only once in our sample. (The results that follow are no different than they would be with those skills included, because the skill similarity weights for these links are, by definition, one.) We are left with 1,933 worker skills and 2,293 job skills. We then use these data to construct supply side and demand side human capital networks as detailed above.

An Endogenous Skill Taxonomy. Placing the skills on a network allows us to use the deep toolbox of network analysis to reveal underlying structure in the job market, such as subpools of labor that are evaluated similarly by employers and categories of jobs for which there is no dedicated labor pool. Here, we partition the human capital networks into groups of related skills using the Louvain method: a standard community-finding algorithm (28). We show that this division is significant using modularity. (*SI Appendix* has additional analysis concerning the significance of this categorization.) The modularity of a partition is proportional to the number of links within a group, relative to what would be expected in a random network. The modularity of the worker network is 0.47, and the modularity of the job network is 0.5, indicating that this division is very strong, and represents real community structure in the network. [It is widely held that any network with modularity above 0.3 has significant community structure (29).] We will call these skill groups "categories."

The skill categories are easily identified using the list of skills in each (*SI Appendix* has the most common skills in each category). The categories are represented by the colors in Fig. 1. For clarity, we have limited the visualization to skills listed by at least 0.5% of worker profiles and at least 0.2% of job listings (the full networks containing all communities are pictured in *SI Appendix*). We have attached names to each of the categories based on that identification. On the worker side, the skill categories are (i) administration, writing, and marketing; (ii) art and design; (iii) software testing; (iv) statistics and mobile development; (v) information technology (IT) administration; and (vi) general programming. Jobs divide into far more categories, presumably because jobs are more specific than workers. The categories here are (i) administrative, (ii) writing, (iii) translation, (iv) marketing, (v) art and design, (vi) music and audio, (vii) software testing, (viii) engineering and physical design, (ix) data handling and statistics, (x) mobile and game development, (xi) IT administration, and (xii) general programming. Note

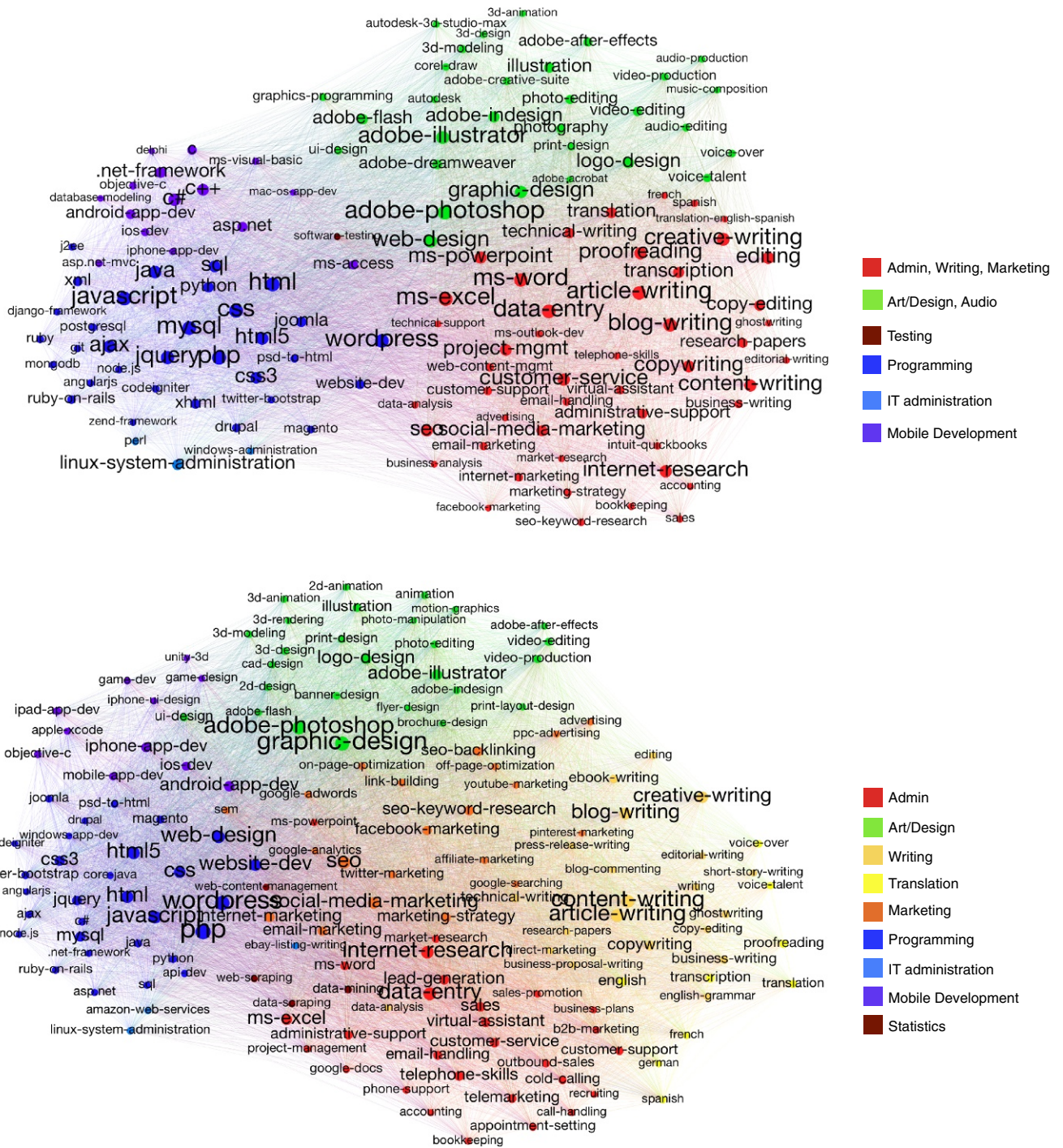


Fig. 1. Worker skill (Upper) and job skill (Lower) human capital networks for the UpWork freelance labor market. For clarity, we have limited the nodes to the most common skills (those held by ~ 0.5% of the workers or listed on ~ 0.2% of job listings). The colors indicate the endogenous categorization of skills derived from the network structure.

that jobs divide into more categories than the workers, suggesting that there are well-defined jobs that lack a well-defined labor pool (e.g., while there are significant numbers of jobs that could be categorized under “translation,” few workers could be definitively identified as “translators”).

This categorization of skills is useful for several reasons. It allows us to quantify the diversity of a worker’s skills. Those whose skills fall into a single category are specialists, while those who bridge categories are generalists. The specialists can be

further divided by their area of specialty. These denote different subpools of labor, which are likely identifiable to employers observing the labor market. Employers use this kind of “low-bandwidth” information as a substitute for more costly search mechanisms (26). In the case of UpWork, there is substantial overlap between the endogenous categories derived from the network and their exogenously generated job categories. (They do not provide information about worker categories.) However, there are some notable differences between the two. The

network suggests that data analysis and software testing require very different sets of skills: the workers in those areas are in distinct labor subpools and are qualified for different jobs. Given that worker search and employer search are guided by these categories, employer/employee matches might be improved by allowing jobs to be categorized by the labor market itself. (*SI Appendix* has more on the differences between the endogenous and exogenous skill taxonomies.)

Human Capital and Wages. We will now illustrate how human capital networks can shed light on the relationship between human capital and wages. We provide three examples where skill interactions might be relevant: worker type, skill diversity, and synergies between skills. By definition, these are aspects of human capital that cannot be captured with independent skills. Note that what follows is not a complete treatment of any of these questions. Our simple illustration faces many of the typical problems that arise in identifying wage determinants, and we will not attempt to make any causal statements. Our intent is to simply illustrate how more nuanced measures of human capital can provide greater insight into the correlations between worker skills and wages. We leave the identification of causal relationships to future work.

Skill Categories. We consider how different types of workers are valued using the skill categories defined above. In the worker network, workers whose skills are in a single area are specialists, representing a worker “type.” Fig. 2*A* shows wages for workers who specialize in each skill category: technical workers earn, on average, \$3.40/h more than artists and designers, who in turn, make an average of \$3.00/h more than writers and administrative workers.

In the job network, workers whose skills are in a single category are qualified for one type of job. Fig. 2*B* shows that there is a similar pattern in wages: workers qualified for tech-

nical jobs make more than those whose employment opportunities are in creative fields, who in turn make more than those who qualify for administrative tasks (a full breakdown of average wages in each worker and job category is in *SI Appendix*). These results are similar to those found in previous studies using UpWork’s exogenous categorization of worker skills (22) and the literature using occupation-level data from O*net and similar sources (17).

Skill Diversity. The human capital networks also provide a method for measuring the diversity of a worker’s skills; a worker whose skills are spread widely on the worker network has a more diverse skill set than one whose skills are tightly clustered in one area. The broader literature on skill diversity is split—some work suggests that workers with more specialized skills make contributions that have higher impact (4), while other work suggests that workers with diverse skills bring more to the problem-solving process (7). Crucially, while most existing work considers diversity in problem-solving and production, it does not consider the wages that workers earn. Here, we measure skill diversity according to the number of categories that a worker’s skills span. Skill sets that are within a single area are specialized, while skill sets spread across multiple areas are more diverse. Here, skill diversity is associated with higher wages—Fig. 2*C* indicates that workers with diverse skill sets earn about \$2.25/h more than those who specialize in a single area.

Skill Synergies. The results of the previous section raise a question: what is it about skill diversity that gives workers higher wages? Looking at the data suggests that workers with diverse skills have one of two potential advantages in the job market. First, skill diversity can expand a worker’s pool of available jobs. For example, a worker with skills in Russian–English translation and mobile application development is unlikely to use her skills in combination. Instead, she uses her diverse skills independently and earns higher wages by virtue of being able to choose from both programming and translation jobs. Second, a worker may have an uncommon combination of skills that can be used synergistically to fill a hole in the market. For example, a worker with user interface design skills and mobile application development skills may use both skills to develop better iOS games. This worker has diverse skills that she uses in combination.

In the data, we can distinguish workers with these two different types of diversity by examining the number of job categories that they cross. Groups of skills that fit into multiple job categories tend to be used independently, while those that fit a single job category tend to be used in combination. We would hypothesize that workers who fill gaps in the market would have higher wages than those who are jacks-of-all-trades. Fig. 2*D* compares workers by their number of worker categories and job categories. As we saw earlier, workers with diverse skills uniformly outperform workers with specialized skills. However, among workers with diverse skills, those who use them in a single job area earn \$1.15/h more than those who use them in two or more job areas. This is consistent with our theory that, among workers with diverse skills, those who use them independently earn more by expanding their range of available jobs but less than those who use a rare combination of skills synergistically.

The Value of Considering Skill Interactions. Finally, we address whether these network-based measures provide insight into wages beyond that provided by individual skills. Our baseline will be the most flexible specification using only individual skills. Specifically, we relate log wages to a vector of skill dummies, where $d_i = 1$ if the worker has that skill and $d_i = 0$ otherwise. We include terms for skills that appear in at least 2% of worker profiles: a total of 62 dummies. (When too many dummies

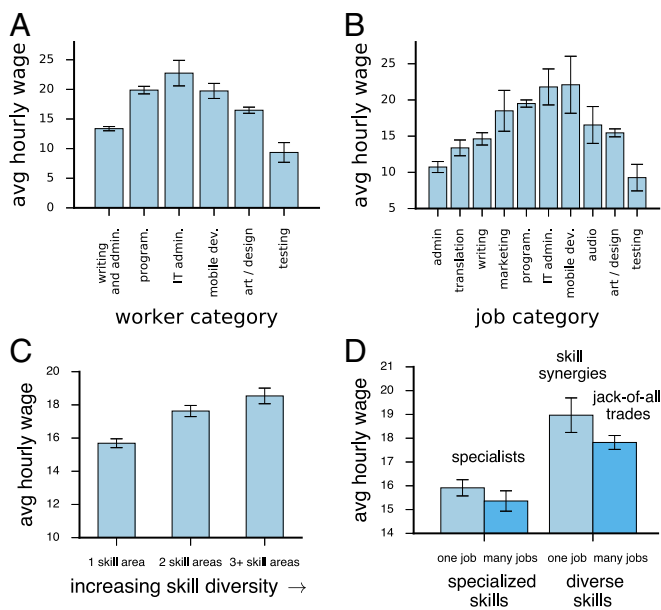


Fig. 2. Average wages for freelance workers with different kinds of human capital. Wages are calculated for all workers who average more than \$1/h. Error bars are a 95% confidence interval on the point estimate. (A) Average wages for workers in different areas. (B) Average wages for workers qualified for jobs in different areas (we have omitted two categories with fewer than 20 workers). (C) Average wages for workers with different levels of skill diversity. (D) Average wages for workers with synergistic skills and those whose skills must be applied independently as a jack-of-all-trades.

are included, the terms become colinear.) As expected, some dummies (e.g., project management) are associated with higher wages, and others (e.g., data entry) are associated with lower wages (Model 1 in *SI Appendix, Table S10*).

We then consider a model that includes both the dummies for individual skills and the network-based worker skill categories: mobile development, testing, programing, IT administration, and art and design (we omit administration as a comparison). The coefficients on these additional terms are significant with signs in the expected directions (Model 2 in *SI Appendix, Table S10*). The effect sizes are quite large: for example, workers in programing fields earn 51% (~\$5.40) more than administrative workers with the same skills, while workers in software testing earn 30% (~\$4.00) less. The adjusted R^2 of this model is higher than that of the baseline model, indicating that the network-based measures explain variance in worker wages, even when controlling for the skills individually.

The story is similar for the other network-based measures. In models containing both the skill dummies and the number of worker categories/job categories crossed, the network-based measures have significant coefficients (Models 3 and 4 in *SI Appendix, Table S10*). The effect sizes on these terms are smaller but still notable. Workers with skills in two different worker categories earn an average of \$0.63 more than their more specialized peers. Workers who qualify for two different types of job earn an average of \$0.62 more than those with more constrained options. Again, the adjusted R^2 of both models is higher than that of the baseline.

Ideally, we would also examine a model with a more traditional one-dimensional measure of human capital, such as years of education or experience. Unfortunately, because of the nature of this dataset, we do not observe either. The closest that we have to such a measure is the number of skills listed on a worker's profile, which is obviously problematic in this context. However, as this type of data becomes more common, we expect that future studies will perform that comparison.

Conclusion

The complex nature of human capital in knowledge-based industry has made the measurement of worker skills increasingly difficult. Placing skills relevant for employment decisions onto a network provides both a way to operationalize the interrelationships between skills and a deep toolbox with which to measure them. As online marketplaces play a greater role in matching workers to employers, human capital networks have an increasing number of practical applications. The detailed skill data in these markets lend themselves to aggregation and algorithmic search (30–32). Even small improvements in these algorithms would reduce search frictions and improve employer–employee matches. The networks can also reveal which skills are complementary to an existing skill set, which would help workers decide on which new skills to acquire and how to best appeal to potential employers.

ACKNOWLEDGMENTS. We thank Brian Kovak, Rebecca Lassem, Sarah Feldt-Muldoon, Seth Richards-Shubik, Ben Tengelsen, Ross O'Connell, and the attendees of the 2016 Networks in Economics Conference.

- Willis RJ (1986) Wage determinants: A survey and reinterpretation of human capital earnings functions. *Handbook of Labor Economics*, eds Ashenfelter O, Card D (North Holland, Amsterdam), Vol 1, pp 525–602.
- Hong L, Page SE (2004) Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proc Natl Acad Sci USA* 101:16385–16389.
- Lazear EP (2005) Entrepreneurship. *J Labor Econ* 23:649–680.
- Adamic LA, et al. (2010) Individual focus and knowledge contribution. arXiv:1002.0561.
- Woolley AW, Chabris CF, Pentland A, Hashmi N, Malone TW (2010) Evidence for a collective intelligence factor in the performance of human groups. *Science* 330:686–688.
- Ren Y, Argote L (2011) Transactive memory systems 1985–2010: An integrative framework of key dimensions, antecedents, and consequences. *Acad Manag Ann* 5:189–229.
- Aggarwal I, Woolley AW (2013) Do you see what I see? The effect of members' cognitive styles on team processes and errors in task execution. *Organ Behav Hum Decis Process* 122:92–99.
- Moreland RL, Levine J, Wingert M (2013) Creating the ideal group: Composition effects at work. *Understanding Group Behavior*, eds Witte EH, Davis JH (Psychology Press, East Sussex, UK), Vol 2, pp 11–35.
- Powell WW, Snellman K (2004) The knowledge economy. *Annu Rev Sociol* 30:199–220.
- Rosen S (1983) Specialization and human capital. *J Labor Econ* 1:43–49.
- Rosen S (1986) The theory of equalizing differences. *Handb Labor Econ* 1:641–692.
- Bowles S, Gintis H, Osborne M (2001) The determinants of earnings: A behavioral approach. *J Econ Lit* 39:1137–1176.
- Heckman JJ, Stixrud J, Urzua S (2006) The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *J Labor Econ* 24:411–482.
- Borghans L, Duckworth AL, Heckman JJ, Ter Weel B (2008) The economics and psychology of personality traits. *J Hum Resour* 43:972–1059.
- Cunha F, Heckman JJ, Schennach SM (2010) Estimating the technology of cognitive and noncognitive skill formation. *Econometrica* 78:883–931.
- Altonji JG (2010) Multiple skills, multiple types of education, and the labor market: A research agenda (September 2010). Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1888514. Accessed October 16, 2017.
- Autor DH, Handel MJ (2013) Putting tasks to the test: Human capital, job tasks, and wages. *J Labor Econ* 31(Suppl 1):S59–S96.
- Jones BF (2014) The human capital stock: A generalized approach. *Am Econ Rev* 104:3752–3777.
- Leung MD (2014) Dilettante or renaissance person? How the order of job experiences affects hiring in an external labor market. *Am Socio Rev* 79:136–158.
- Allinson CW, Hayes J (1996) The cognitive style index: A measure of intuition-analysis for organizational research. *J Manag Stud* 33:119–135.
- Altonji JG, Kahn LB, Speer JD (2014) Trends in earnings differentials across college majors and the changing task composition of jobs. *Am Econ Rev* 104:387–393.
- Agrawal A, Horton J, Lacetera N, Lyons E (2015) Digitization and the contract labor market: A research agenda. *Economic Analysis of the Digital Economy*, eds Goldfarb A, Greenstein SM, Tucker CE (Univ of Chicago, Chicago), pp 219–250.
- Horton JJ (2010) Online labor markets. *Proceedings of the International Workshop on Internet and Network Economics* (Springer, Berlin), pp 515–522.
- Agrawal AK, Lacetera N, Lyons E (2013) Does information help or hinder job applicants from less developed countries in online markets? Available at www.nber.org/papers/w18720. Accessed October 22, 2017.
- Ghani E, Kerr WR, Stanton C (2014) Diasporas and outsourcing: Evidence from odesk and India. *Manag Sci* 60:1677–1697.
- Barach MA (2015) Search, screening, and information provision: Personnel decisions in an online labor market. Available at moshebarach.com/files/JMP.pdf. Accessed May 7, 2017.
- Chen DL, Horton JJ (2016) Research note—Are online labor markets spot markets for tasks? A field experiment on the behavioral response to wage cuts. *Inf Syst Res* 27:403–423.
- Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech Theory Exp* 2008:P10008.
- Newman ME, Girvan M (2004) Finding and evaluating community structure in networks. *Phys Rev E* 69:026113.
- Autor DH (2001) Wiring the labor market. *The J Econ Perspect* 15:25–40.
- Agrawal A, Lacetera N, Lyons E (2016) Does standardized information in online markets disproportionately benefit job applicants from less developed countries? *J Int Econ* 103:1–12.
- Horton JJ (2015) The effects of algorithmic labor market recommendations: Evidence from a field experiment. *J Labor Econ* 35:345–385.